



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/297,038	08/24/1999		6715/56610	9199

7590 04/07/2004

JAY H MAIOLI
COOPER & DUNHAM
1185 AVENUE OF THE AMERICAS
NEW YORK, NY 10036

EXAMINER

KNEPPER, DAVID D

ART UNIT	PAPER NUMBER
----------	--------------

2654

DATE MAILED: 04/07/2004

16

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/297,038

Applicant(s)

SEYA

Examiner

David D. Knepper

Art Unit

2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 07 January 2004.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-12 and 24-30 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-12 and 24-30 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

1. Applicant's correspondence filed on 7 January 2004 (paper #15) has been received and considered. Claims 1-12 and 24-30 are pending. Claims 13-23 and 31-52 have been canceled.

Title

2. The title is objected to because it fails to provide a useful description of the invention.

Abstract

3. The Abstract of the Disclosure is objected to because it repeats information in the title and is not commensurate in scope with the invention as described in the specification. The abstract indicates that that the invention will separate musical, vocal input into language (presumably, the lyrics) and accompaniment information (presumably, the musical score to include vocalists and/or any other music related data). The invention then purports to translate the vocal (lyrics) and produce a second vocal output, which is a translated version of the input.

The applicant's arguments on page 14 of paper 15 indicate that the applicant believes "separating speech from musical accompaniment is well understood in the art." However, this is a primary element of the claims and applicant never provided evidence to support this seeming admission.

Correction is required. See M.P.E.P. § 608.01(b).

Drawings

4. The drawings are objected to under 37 CFR 1.83(a). The drawings must show every feature of the invention specified in the claims. Therefore, the information that is being

processed must be shown or the feature(s) canceled from the claim(s). For example, the “vocal information”, “accompaniment”, “language lyric” and “musical information.” The drawings fail to show how any type of data separation is performed. The drawings must show the data input relied upon as well as the method for processing the data to achieve the desired result commensurate with the description and claims.

No new matter should be entered.

The arguments on page 13 of paper #15 point out a list of desired results (functions) but no support to indicate how or what processing is done regarding input and output relationships.

Figures 4-6 only show a desired result without providing any useful showing of how the information is determined or extracted. The information itself is NOT illustrated. Only text is provided which is only sufficient if the information and methods (or means plus function) for analyzing, canceling and extracting the information is obvious or well known per se.

New Matter

5. The objection to new matter is overcome.

Claims

6. Claims 1-12 and 24-30 are rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claims 1 and 24 are rejected as noted below.

The claim language limits the input to “first vocal-containing musical number

information". The claims are confusing because they contradict the specification.

"Generating the first language lyric information by speech recognition of the first vocal information part" indicates that someone speaks the lyrics into the device. From the antecedent reference of the separation unit, the lyrics must be derived from speech. However, the accompaniment is not limited to any particular "musical information", be it vocal, instrumental or otherwise. This contradicts the specification such as page 23, which indicates that some unique form of karaoke information is required as a substitute for standard stereo audio such as left/right (L/R) and center information that is normally used for some types of stereo coding. Confusion exists because the claims and the specification fail to provide reasonable antecedent basis to form a meaningful relationship for interpretation of the claims. The claims do not indicate the need for any particular karaoke information and must therefore apply to generic (monophonic) audio input.

The claims are broad enough to include the separation of any speech from any input audio. For example, this would include choral music and the separation of a typical 4-part (SATB) harmony into the desired lyrics of one or more parts (the lyrics could vary among parts). Other possibilities would include one or more singers accompanied by a band or orchestra in which separation of parts could be much more complicated between singers, instruments and/or other accompaniment parts. However, the "vocal separation unit 212" is described on page 23 as including "vocal canceling unit 212a" which contains a digital filter capable of erasing the vocal part. It is unclear what relationship between "vocal information", "accompaniment information" and "musical information" is established and how the relationships should be defined and/or separated.

It is unclear whether the claimed "language lyric information" is part of the "vocal information D3 or the karaoke information D2 (page 23, line 18). The combination claimed fails to provide an antecedent basis for a clear relationship to the specification.

It is unclear whether the applicant intends to limit the invention to musical related information or whether the combination of data can include a wider variety of multimedia information. Therefore, the separation of data is interpreted to be broad enough to include various types of information known in the art.

The only specific application mentioned in the specification is for karaoke related data. However, neither the specification nor the claims clearly describe any particular requirements for data structure or data format for karaoke devices and/or methods for using karaoke devices.

7. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

8. Claims 1-12 and 24-30 are rejected under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention.

The claims (i.e. – 1 and 24) are drawn towards "an information processing apparatus". There is no particular limitation to the application of the apparatus itself. The only limitations are functional. The primary limitations are the ability to accept combinations of vocal, accompaniment and musical information and to separate them. Once they are separated, the

claims indicate that the vocal information portion will be further analyzed by speech recognition, which will be used to generate lyrics and to translate lyrics from a first language to a second language. Then the second language translation is synthesized with the accompaniment. "Accompaniment" would not be limited to non-speech. Therefore, a cappella singing where the melody and accompaniment are separated require support as well as synthesizing speech to form the accompaniment.

The "separation" is described on page 23 as requiring the input data to be limited such that it can only work while "canceling the speech signals fixed at the center on stereo reproduction..." However, the claims are not so limited nor does the specification explain why one of ordinary skill in the art would expect generic "data" as claimed to be so limited.

The specification on page 6 indicates that "the required information" is not particularly limited but may include various data "such as audio information, text information, image information or the picture information as later explained..." The specification fails to limit the information to any particular format or combination of data. This makes it unlikely that one of ordinary skill in the art could hope to predict how to process the data since there is no way to know what needs to be processed. The desired result of the processed data is similarly vague.

On page 12, last paragraph, the server device is described in part as containing "an assessment processing unit 105 for assessment processing for the user and an interfacing unit 106 for having communication with the intermediate transmission device 2." The function of the "assessment processing unit 105" is undefined. Neither description of the data being assessed nor any description of the resultant assessment is provided to give life and meaning to these terms.

Applicant reliance upon foreign [Japanese] documents H-3-139923 or 3-13922 for teaching how to make and use TwinVQ (page 14) is acceptable since the statements indicate that the form of data compression is not considered inventive. However, since the data and the form of compression have no bounds, the burden on the applicant to provide a specific form of signal processing that can apply to any audio input, regardless of the type or format is that much more difficult to meet.

The original specification, page 14, second full paragraph failed to indicate what data is “collated”. Supposedly, “The terminal ID data of the portable terminal device 3” is magically collated with “the terminal ID data of the portable terminal device that is currently able to use the information distribution system”. How is a single device “currently able to use the information...” identified? Since page 13 implies use of the Internet, why would only one device (such as a computer) be able to use such information? The Examiner cannot resolve how “collation processing” is used (i.e. –“the results of collation”) to decide what is “permitted”.

The additional relationship indicating a use fee is not considered consistent with page 15, lines 9-18, which refers to “a charging circuit, for supplying the power to the various parts.” This provided an antecedent for the term “electrical charging” defined as something that supplies electrical power rather than charging fees as the new language would imply. The following paragraph was the original complaint under 35 US 112, first paragraph of the original language.

The last paragraph of page 14 (continuing to page 15) does not explain what sort of “assessment” is desired. The sentence “... the assessment processing unit 105 performs the processing of assessment of the amount in meeting with the state of use of the information distribution system by the user...” is incoherent. Neither the data being assessed, the process by

which assessment is performed, what “amount” (amount of what?) is utilized nor the “state of use” are defined. The applicant throws about these terms without being given any meaningful definitions thereof. The example given is nonsense: “the request information for information copying or electrical charging...” How can a request for “information copying” be treated as an alternative to “electrical charging” (charging a battery?). What does each of these things really mean?

No evidence that the applicant’s invention can take generic information and separate portions of vocal information and accompaniment (vocal and/or instrumental) is presented by the applicant.

Page 18, second paragraph mentions “speech recognition translation unit 321” and “speech synthesis unit 322” but provides no details capable of actually achieving the desired results of either unit.

Page 19, last paragraph, indicates that “speech recognition translation unit 321 is fed with the vocal information transmitted along with the karaoke information after separation by the vocal separation unit 212 of the intermediate transmission device 2, and performs speech recognition of the vocal information.” Again, no details for separating desired audio (a particular vocal) from other audio data is even offered by the applicant. Similarly, no details for performing speech recognition and translation to another language are offered. As a minimum, the drawings should show the steps of analysis necessary to input typical song data and extract specific parameters that can be analyzed by a computer to determine the desired results. Details must be provided giving a reasonably detailed explanation of how one of ordinary skill in the art could expect successful separation, recognition and translation.

Paper #9 presents arguments on page 7 that page 23 of the specification provides support for a separation unit. However, page 23 of the specification is also characterized as a non-limiting example. Reading the specification itself, it is found that page 23 explicitly states that "the detailed structure of the vocal canceling unit 212a is omitted." It goes on to say that "...the vocal canceling unit 212a generates the karaoke information D2 using the well-known technique of canceling the speech signals fixed at the center on stereo reproduction with the {(L channel data)-(R channel data)}." However, the applicant has never limited the vocal information to a particular type of stereo data. This technique would not necessarily reside in a digital filter nor would a generic digital filter have the ability to isolate specific vocal information from other types of vocal and/or other audio information. Therefore, the applicant is required to show a technique for separating speech from other audio data that is commensurate in scope with the claims and the specification. At the very least, a specific digital filter capable of such functionality must be shown. The applicant is reminded of the specification on page 6 (noted above) which indicates that "the required information" is not particularly limited but may include various data "such as audio information, text information, image information or the picture information as later explained..."

Page 20, first full paragraph, indicates that "speech synthesis unit 322 first generates the novel vocal information (audio data) sung with the lyric of the as-translated second language, based on the second language lyric information generated by the speech recognition translation unit 321." No details for performing such a desired manner of synthesis are provided. The apparatus and method for analysis as well as the parameters for modeling "original vocal information" must be provided. Similar details for synthesizing speech with musical properties

must be shown that will allow utilization of "original vocal information." Further details are necessary to show "original vocal information" specifics with regard to music and speech. Such details must include time and frequency as it relates to musical pitch and vocal tract and/or other information that is specific to language idiosyncrasies. For example, in English, changes in pitch do not change the literally meaning of a word, but in certain Eastern languages (i.e. - Chinese) a rising or falling pitch could change the meaning of an otherwise identical pronunciation. Such details provide interesting challenges to the desired results of applicant's invention. However, no details are provided to address these or even more basic information.

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

10. Claims 1-12 and 24-30 are rejected under 35 U.S.C. § 103 as being unpatentable over Stelovsky (5,613,909) in view of Bordeaux (4,852,170) and Lyberg (5,546,500).

As per claims 1 and 24, "information processing" is taught by both references:

"separating a first vocal information part in a first language and an

accompaniment information part . . . generating first language lyric information” (Stelovsky teaches the separation of lyrics in figure 5 – see also column 9, lines 12-21 where he indicates the lyrics can be viewed separately from the music and accompanying video by using separate tracks that are common in karaoke applications);

“translating the generated first language letter information into the second language letter information” (suggested in column 14, lines 21-22 using direct translation into another language); and

“synthesizing the second language lyric information” (suggested in column 14, lines 18-19 where he teaches that the audio track can be generated rather than recorded (e.g. using a speech generator.) – see also col. 14, lines 22-24 that teaches Any of the tracks of presentation can be generated remotely and transmitted using any existing communication means. Thus, it would have been obvious to combine or otherwise synthesize any known combinations of the data).

It is noted that Stelovsky does not explicitly teach the use of “speech recognition” to perform translation. However, he teaches that translation is obvious in combination with a karaoke or other multimedia separation of data elements in order to facilitate education and/or entertainment. Bordeaux and Lyberg teach details for performing speech recognition and in column 12, lines 60-65, Bordeaux teaches that for use in foreign languages . . . a different natural language or orthographic translator would be employed. It would have been obvious for a person having ordinary skill in the pertinent art, at the time the invention was made, to combine a speech recognition based translator such as Bordeaux with the device of Stelovsky because Stelovsky specifically invites the use of future facilities (col. 14, line 11) which include

translation into other languages as noted above. Lyberg explicitly recites Bordeaux in column 1, line 24 and is utilized because he clearly teaches that it is known to combine synthesis with a translation device (see abstract) in such a way as to preserve prosodic information even after translation to further improve translation. Thus, Lyberg teaches that a combination of speech recognition for translation to a second language can be used to preserve stresses in the first language (abstract).

Claims 2-12 and 25-30 are rejected under similar arguments as presented above. Although the claims are unclear, it is presumed that the applicant is attempting to limit some of the synthesis related elements to preserving information gathered during analysis or recognition. This is taught by Lyberg who preserves prosody following recognition and translation for use in synthesis.

Bordeaux teaches details regarding the identification of phoneme strings (words) to be translated into any natural language. Lyberg teaches that such translation devices can separate the language and prosody in order to allow preservation of the prosody after translation.

Remarks

11. The applicant's remarks make generic statements that changes were made in response to the rejections. However, it is not seen how any of the changes actually overcome the deficiencies.

The objection under 37 CFR 1.83 argued by the applicant on page 16 of paper #15 is an attempt to derive clarity from the specification. Figures showing how the most significant components of the invention work should be shown.

The arguments on pages 16-17 of paper 15 regarding the vague use of an “ID” and “collation” arise because the specification was so poorly written that it is unclear why portions of the specification seem to contradict each other as the applicant submits more and more modifications thereto. If the original application was so poorly written, then the application should be completely re-written (and possibly re-filed as a continuation-in-part to preserve common subject matter).

The argument on page 18 of paper #15 that the processing of “non-lyrical music” will not change how the music is treated is contrary to the claims. The claimed invention would be inoperative without any vocal information to separate, process and synthesize.

The arguments on pages 19 of paper #15 about stereo channels would not enable the claims because they perform no processing of audio in a stereo format.

The arguments on pages 19-20 of paper #15 about the Examiner’s example of music related data that is not supported by the specification misses the point. The applicant does not seem to care whether the claims accurately depict the invention. None of the claims even mention the term “karaoke” nor do they imply particular information that exclude any known music that includes speech. The applicant inexplicably argues that “the intricacies of speech recognition, speech separation and speech canceling are not the subject of the current invention” even though this is the focus of the subject matter appearing in the claims.

The applicant’s argument on pages 20-21 of paper #15 that the data could include video or picture data is immaterial since this is not claimed nor is any processing of video or picture data enabled by the specification.

The applicant’s argument on pages 21-22 regarding the art of “karaoke” is immaterial

since this is not claimed. This is one point that the Examiner has been trying to emphasize but the applicant does not seem to care.

The applicant's arguments on pages 26-28 repeat previous arguments that the primary elements of the claims are well known to those of ordinary skill in the art. To the contrary, "separation" of desired signals, "speech recognition", "synthesis" and "translation" are all cutting edge technologies that many spend years of research developing. To make them work together would require a great deal of expertise in all four fields of endeavor. Failure to know which methodology used in each field to achieve the desired result could require years of research to develop a new technique that would enable one or another of each field of endeavor to properly process the input from one to the other to properly convert each type of data that might come into each of the "technologies" claimed.

The applicant's arguments against the prior art on pages 29-32 fails to consider that the claims do not limit the type of input data other than to require "vocal" and "musical" data to be included. If the applicant's previous arguments that one of ordinary skill in the art would be familiar with the type of data commonly used in karaoke, then Stelovsky is considered evidence that one of ordinary skill in the art would find it obvious to separate the data that is commonly stored together in separate tracks. It is the applicant's decision as to whether the claimed data should remain broad enough to cover any combination of vocal, music, etc. or whether it must be narrowed to overcome the prior art.

These arguments against the prior art does not address the fact that Stelovsky shows that his device can separate lyrics from musical information for use in a karaoke device. This is significant because the applicant indicates in the specification that the invention is related to a

karaoke device but the claims are framed in a much broader sense to cover other uses of audio.

Stelovsky suggests translation into another language in column 14, lines 21-22 and both Lyberg and Bordeaux show that translation can be done using speech recognition. Stelovsky also teaches that his invention is designed to be flexible to incorporate future improvements. Specifically, he is able to allow programming to utilize recorded or generated audio to include a speech generator (col. 14, lines 11-20). Thus, the combination of references teaches that it is obvious to perform translation using text and/or speech recognition to improve speech comprehension and improve the usefulness (Industrial Applicability) of Stelovsky's system.

12. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

13. **Any response to this action should be mailed to:**

Box AF
Commissioner of Patents and Trademarks

Washington, D.C. 20231

or faxed to:

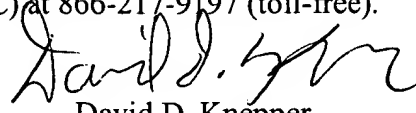
TC2600 Fax Center
(703) 872-9315

Hand-delivered responses should be brought to Crystal Park II, 2121 Crystal Drive, Arlington, VA., Sixth Floor (Receptionist).

14. Any inquiry concerning this communication or earlier communications from the examiner should be directed to David D. Knepper whose telephone number is (703) 305-9644. The examiner can normally be reached on Mon-Thursday 7:30 a.m. - 6:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (703) 305-9645. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).


David D. Knepper
Primary Examiner
Art Unit 2654